STA 6207 – Exam 2 – Fall 2018 **PRINT** Name ANSWER KEY

Conduct all tests at $\alpha = 0.05$ significance level

All Questions are based on the following 2 regression models, where SIMPLE REGRESSION refers to the case where p=1, and X is of full column rank (no linear dependencies among predictor variables).

$$\text{Model1}: Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \varepsilon_i \quad i = 1,\ldots,n \quad \varepsilon_i \sim NID(0,\sigma^2)$$

$$\text{Model2}: \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \mathbf{X} \equiv n \times p' \quad \boldsymbol{\beta} \equiv p' \times 1 \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

Given: $\dfrac{d(\mathbf{a}'\mathbf{x})}{d\mathbf{x}} = \mathbf{a} \qquad \dfrac{d(\mathbf{x}'\mathbf{A}\mathbf{x})}{d\mathbf{x}} = 2\mathbf{A}\mathbf{x}$ (A symmetric) $\qquad E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = tr(\mathbf{A}\mathbf{V}_\mathbf{Y}) + \boldsymbol{\mu}_\mathbf{Y}'\mathbf{A}\boldsymbol{\mu}_\mathbf{Y}$

Cochran's Theorem: Suppose **Y** is distributed as follows with nonsingular matrix **V**:

$\mathbf{Y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{V}) \qquad r(\mathbf{V}) = n$ then if **AV** is idempotent:

1. $\mathbf{Y}'\left(\dfrac{1}{\sigma^2}\mathbf{A}\right)\mathbf{Y}$ is distributed non-central χ^2 with: (a) df $= r(\mathbf{A})$ and (b) Noncentrality parameter: $\Omega = \dfrac{1}{2\sigma^2}\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$

2. $\mathbf{Y}'\mathbf{A}\mathbf{Y}, \quad \mathbf{Y}'\mathbf{B}\mathbf{Y}$ independent if $\mathbf{AVB} = \mathbf{0}$

Critical Values for t and F-distributions

F-distributions indexed by numerator df across top of table

df	t(.05)	t(.025)	X2(.05)	F(.05,1)	F(.05,2)	F(.05,3)	F(.05,4)	F(.05,5)	F(.05,6)
1	6.314	12.706	3.841	161.448	199.500	215.707	224.583	230.162	233.986
2	2.920	4.303	5.991	18.513	19.000	19.164	19.247	19.296	19.330
3	2.353	3.182	7.815	10.128	9.552	9.277	9.117	9.013	8.941
4	2.132	2.776	9.488	7.709	6.944	6.591	6.388	6.256	6.163
5	2.015	2.571	11.070	6.608	5.786	5.409	5.192	5.050	4.950
6	1.943	2.447	12.592	5.987	5.143	4.757	4.534	4.387	4.284
7	1.895	2.365	14.067	5.591	4.737	4.347	4.120	3.972	3.866
8	1.860	2.306	15.507	5.318	4.459	4.066	3.838	3.687	3.581
9	1.833	2.262	16.919	5.117	4.256	3.863	3.633	3.482	3.374
10	1.812	2.228	18.307	4.965	4.103	3.708	3.478	3.326	3.217
11	1.796	2.201	19.675	4.844	3.982	3.587	3.357	3.204	3.095
12	1.782	2.179	21.026	4.747	3.885	3.490	3.259	3.106	2.996
13	1.771	2.160	22.362	4.667	3.806	3.411	3.179	3.025	2.915
14	1.761	2.145	23.685	4.600	3.739	3.344	3.112	2.958	2.848
15	1.753	2.131	24.996	4.543	3.682	3.287	3.056	2.901	2.790
16	1.746	2.120	26.296	4.494	3.634	3.239	3.007	2.852	2.741
17	1.740	2.110	27.587	4.451	3.592	3.197	2.965	2.810	2.699
18	1.734	2.101	28.869	4.414	3.555	3.160	2.928	2.773	2.661
19	1.729	2.093	30.144	4.381	3.522	3.127	2.895	2.740	2.628
20	1.725	2.086	31.410	4.351	3.493	3.098	2.866	2.711	2.599
30	1.697	2.042	43.773	4.171	3.316	2.922	2.690	2.534	2.421
40	1.684	2.021	55.758	4.085	3.232	2.839	2.606	2.449	2.336
50	1.676	2.009	67.505	4.034	3.183	2.790	2.557	2.400	2.286
60	1.671	2.000	79.082	4.001	3.150	2.758	2.525	2.368	2.254
70	1.667	1.994	90.531	3.978	3.128	2.736	2.503	2.346	2.231
80	1.664	1.990	101.879	3.960	3.111	2.719	2.486	2.329	2.214
90	1.662	1.987	113.145	3.947	3.098	2.706	2.473	2.316	2.201
100	1.660	1.984	124.342	3.936	3.087	2.696	2.463	2.305	2.191
110	1.659	1.982	135.480	3.927	3.079	2.687	2.454	2.297	2.182
120	1.658	1.980	146.567	3.920	3.072	2.680	2.447	2.290	2.175
130	1.657	1.978	157.610	3.914	3.066	2.674	2.441	2.284	2.169
140	1.656	1.977	168.613	3.909	3.061	2.669	2.436	2.279	2.164
150	1.655	1.976	179.581	3.904	3.056	2.665	2.432	2.274	2.160

Q.1. A study considered noise level of the Teheran-Karaj express train (Y, in dB) in terms of distance to the center of the track (X_1, in meters) and speed of the train (X_2, in km/h), with n = 50. Consider the following models (in matrix form).

Model 0: $E\{Y\} = \beta_0$ Model 01: $E\{Y\} = \beta_0 + \beta_1 X_1$ Model 02: $E\{Y\} = \beta_0 + \beta_2 X_2$

Model 012: $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ $\mathbf{Y'Y} = 348654$

p.1.a. The sum of the speeds of the 50 observations is $\sum_{i=1}^{50} Y_i = 4174.2$. For model 0, obtain: $(X_0'X_0)^{-1} = \frac{1}{50} = 0.02$

$\mathbf{X_0'X_0}$, $\mathbf{X_0'Y}$, $(\mathbf{X_0'X_0})^{-1}$, $\hat{\beta}_0$, $\mathbf{Y'P_0Y}$ $\underset{\sim}{X_0} = \underset{\sim}{1}_{50}$ $X_0'X_0 = 50$ $X_0'Y = \sum Y_i = 4174.2$

$\hat{\beta}_0 = \frac{1}{50}(4174.2) = 83.484$ $\underset{\sim}{Y}'P_0\underset{\sim}{Y} = \hat{\beta}_0 X_0'Y = 83.484(4174.2)$
$= 348478.9$

4 each

$\mathbf{X_0'X_0} =$ 50 $\mathbf{X_0'Y} =$ 4174.2 $(\mathbf{X_0'X_0})^{-1} = \frac{1}{50}$ $\hat{\beta}_0 =$ 83.484 $\mathbf{Y'P_0Y} =$ 348478.9

p.1.b. For Models 01, 02, and 012, you obtain the following

$\mathbf{X_*'Y}$, $\hat{\beta}_*$ Compute $\mathbf{Y'P_{01}Y}$, $\mathbf{Y'P_{02}Y}$, $\mathbf{Y'P_{012}Y}$ and MSE_{012}

X01'Y	Beta-hat01		X02'Y	Beta-hat02		X012'Y	Beta-hat012
4174.2	88.5825		4174.2	75.7838		4174.2	80.1494
186706	-0.1133		332604.84	0.0967		186706	-0.1158
						332604.84	0.1073

$\underset{\sim}{Y}'P_{01}\underset{\sim}{Y} = 4174.2(88.5825) + 186706(-.1133) = 348607.3$

$\underset{\sim}{Y}'P_{02}\underset{\sim}{Y} = 4174.2(75.7838) + 332604.84(.0967) = 348499.6$

$\underset{\sim}{Y}'P_{012}\underset{\sim}{Y} = 4174.2(80.1494) + 186706(-.1158) + 332604.84(.1073) = 348627.6$

$SSE_{012} = \underset{\sim}{Y}'(I - P_{012})\underset{\sim}{Y} = 348654 - 348627.6 = 26.4$ $MSE_{012} = \frac{26.4}{50-3} = 0.562$

4 each

$\mathbf{Y'P_{01}Y} =$ 348607.3 $\mathbf{Y'P_{02}Y} =$ 348499.6 $\mathbf{Y'P_{012}Y} =$ 348627.6 $MSE_{012} =$ 0.562

p.1.c. Obtain the Sequential and Partial sums of squares for X_1 and X_2, and their corresponding F-statistics.

3 each All R's begin w/ 348 2 each 3 each 2 each

Variable	Sequential SS	Sequential F		Partial SS	Partial F
X1	128.4	228.5		128.0	227.8
X2	20.3	36.1		20.3	36.1

$R(\beta_1|\beta_0) = 607.3 - 478.9 = 128.4$ $R(\beta_1|\beta_0, \beta_2) = 627.6 - 499.6 = 128.0$

$R(\beta_2|\beta_0, \beta_1) = 627.6 - 607.3 = 20.3$ $R(\beta_2|\beta_0, \beta_1) = 20.3$

Q.2. Consider the general linear test $H_0 : \mathbf{K'\beta} = \mathbf{0}$ where **K'** has q ≤ p' linearly independent rows.

p.2.a. Derive the mean vector and variance-covariance matrix of $\mathbf{K'\hat{\beta}}$.

$E\{\hat{\beta}\} = \beta \quad V\{\hat{\beta}\} = \sigma^2(X'X)^{-1}$

$\Rightarrow E\{K'\hat{\beta}\} = K'E\{\hat{\beta}\} = K'\beta$ 8

$V\{K'\hat{\beta}\} = K'\sigma^2(X'X)^{-1}K = \sigma^2(K'(X'X)^{-1}K)$ 8

p.2.b. Show that $Q = \left(\mathbf{K'\hat{\beta}}\right)'\left[\mathbf{K'(X'X)^{-1}K}\right]^{-1}\mathbf{K'\hat{\beta}}$ and SSE are independent. Hint: Write $Q = \mathbf{Y'AY}$.

16

$SSE = Y'(I-P)Y \qquad V\{Y\} = \sigma^2 I \Rightarrow V = I$

$Q = \hat{\beta}'K[K'(X'X)^{-1}K]^{-1}K'\hat{\beta} = Y'X(X'X)^{-1}K[K'(X'X)^{-1}K]^{-1}K'(X'X)^{-1}X'Y$

$(I-P)(I-P) = I-2P+PP = I-2P+P = I-P$ ✓

$AVAV = X(X'X)^{-1}K[K'(X'X)^{-1}K]^{-1}K'\underbrace{(X'X)^{-1}X'X(X'X)^{-1}}_{(X'X)^{-1}}K[K'(X'X)^{-1}K]^{-1}K'(X'X)^{-1}X'$

I

$= X(X'X)^{-1}K[K'(X'X)^{-1}K]^{-1}K'(X'X)^{-1}X' = AV$ ✓

$AVB = X(X'X)^{-1}K[K'(X'X)^{-1}K]^{-1}K'(X'X)^{-1}X'(I-P)$

$B \downarrow I-P$

32

$X'(I-P) = X' - X'X(X'X)^{-1}X' = X' - X' = 0$ ✓

Independent

Q.3. A study was conducted, relating an abrasivity index measure (Y) to p = 4 predictors: UCS (X_1), BTS (X_2), and two brittleness indices: B_1 (X_3) and B_3 (X_4) in a sample of igneous rocks. The model fit is given below along with computations.

$n = 40 \quad p' = 5 \quad n - p' = 35$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i \qquad \varepsilon \sim NID\left(0, \sigma^2\right)$$

X'X						X'Y
40	102.81	51.975	62.766	93.597		109.60
102.81	322.0451	121.9227	156.505	223.8523		263.04
51.975	121.9227	80.71724	93.99107	136.4427		152.32
62.766	156.505	93.99107	122.2703	153.0712		186.27
93.597	223.8523	136.4427	153.0712	241.7843		261.22
INV(X'X)						Beta-hat
1.509	-0.071	1.554	-0.634	-0.994		5.671
-0.071	0.023	0.045	-0.017	-0.008		-0.255
1.554	0.045	3.028	-1.150	-1.624		4.972
-0.634	-0.017	-1.150	0.482	0.605		-1.305
-0.994	-0.008	-1.624	0.605	0.930		-2.859
Y'Y	Y'(J/n)Y	Y'PY				
331.28	300.30	322.05				

p.3.a. Compute SSE, df_E and MSE

$SSE = 331.28 - 322.05 = 9.23$ 4

$df_E = 35$ 2

$MSE = \frac{9.23}{35} = 0.264$ 2

p.3.b. Compute SSR, df_R and MSR

$SSR = 322.05 - 300.30 = 21.75$ 4

$df_R = p = 4$ 2

$MSR = \frac{21.75}{4} = 5.438$ 2

p.3.c. Test H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (abrasivity index is not associated with any of the predictors)

$F_{obs} = \frac{MSR}{MSE} = \frac{5.438}{0.264} = 20.60 \qquad F_{.05,4,35} \approx 2.648$

4 3 2

Test Statistic $F = 20.60$ Rejection Region $F \geq 2.648$ P > < .05

p.3.c. Test H_0: $\beta_3 = \beta_4$ (The coefficients for the two brittleness indices are equal)

K' = $[0\ 0\ 0\ 1\ -1]$ 4 **m** = $[0]$ 2 $K'\hat{\beta} = 1(-1.305) + (-1)(-2.859) = 1.554$

$$K'(X'X)^{-1}K = \begin{bmatrix}0.360 & -.009 & 0.474 & -0.123 & -0.325\end{bmatrix}\begin{bmatrix}0\\0\\0\\1\\-1\end{bmatrix} = 0.202$$

$q = 1 \qquad s^2 = 0.264$

$$Q = (1.554)^2 [0.202]^{-1} = 11.96$$

$$F = \frac{Q/1}{0.264} = \frac{11.96}{0.264} = 45.3 \qquad F_{.05,1,35} \approx 4.128$$

10 3 2

Test Statistic $F = 45.3$ Rejection Region $F \geq 4.128$ P > < .05

46

Q.4. An experiment was conducted relating springiness in berries (Y, ~~in mm~~) to sugar equivalent (X, in g/L). There were $c = 4$ distinct sugar equivalent groups, with $n_j = 5$ berries per group. The lack-of-fit test for a linear relation is:

$$H_0: E\{Y_{ij}\} = \mu_j = \beta_0 + \beta_1 X_j \quad i = 1,\ldots,5;\ j = 1,\ldots,4 \quad H_A: E\{Y_{ij}\} = \mu_j \neq \beta_0 + \beta_1 X_j$$

ANOVA												
	df	SS	MS	F	gnificance F	j	X_j	n_j	Yhat_j	Ybar_j	s^2_j	$\bar{y}_j - \hat{y}_j$
Regressio	1	245.74	245.74	57.40	0.0000	1	176.5	5	21.89	21.77	0.38	-0.12
Residual	18	77.06	4.28			2	209.3	5	18.04	18.77	2.90	0.73
Total	19	322.80				3	225	5	16.20	15.41	11.30	-0.79
						4	259.5	5	12.14	12.32	3.18	0.18
	Coefficients	tandard Err	t Stat	P-value								0
Intercept	42.6025	3.4019	12.5233	0.0000								
sugCont	-0.1174	0.0155	-7.5763	0.0000								

Note: $s_j^2 = \dfrac{\sum_{i=1}^{n_j}\left(Y_{ij} - \bar{Y}_j\right)^2}{n_j - 1}$ $n_j > 1$, 0 otherwise

$n = 20$ $\quad C = 4$

p.4.a. Give the fitted value for the linear regression for the 4th group ($X_4 = 259.5$).

3

$$\hat{Y}_{14} = 42.6025 - .1174(259.5) = 12.14$$

p.4.b. Compute the Pure Error Sum of Squares, degrees of freedom and Mean Square.

8

$$SSPE = \sum_{j=1}^{4} (n_j - 1)\, s_j^2 = 4[0.38 + 2.90 + 11.30 + 3.18] = 4(17.76) = 71.04$$

$$df_{PE} = n - c = 20 - 4 = 16 \qquad MSPE = \frac{71.04}{16} = 4.44$$

3 $\quad$ 2

SS_{PE} = 71.04 $\quad$ df_{PE} = 16 $\quad$ MS_{PE} = 4.44

p.4.c. Compute the Lack-of-Fit Sum of Squares, degrees of freedom and Mean Square.

$$SSLF = \sum_{j=1}^{4} n_j (\bar{y}_j - \hat{y}_j)^2 = 5\left((-0.12)^2 + 0.73^2 + (-0.79)^2 + 0.18^2\right)$$

$$= 5(1.2038) = 6.02 \qquad df_{LF} = c - 2 = 4 - 2 = 2$$

8

$$MSLF = \frac{6.02}{2} = 3.01$$

2 $\quad$ 2

SS_{LF} = 6.02 $\quad$ df_{LF} = 2 $\quad$ MS_{LF} = 3.01

p.4.d. Give the Test Statistic, Rejection Region, and P-value relative to .05 for the Lack-of-Fit test.

$$F = \frac{MSLF}{MSPE} = \frac{3.01}{4.44} = 0.68 \qquad F_{.05,2,16} = 3.634$$

4 $\quad$ 3 $\quad$ 2

Test Statistic: $F_{LF} = 0.68$ $\quad$ Rejection Region: $F_{LF} \geq 3.634$ $\quad$ P-value > or < .05

36

Q.5. Regression models were fit, relating various crime rates for U.S. states to a set of 25 predictors. The researchers fit the full model with all 25 predictors (say $X_1,\ldots,X_{25}$) and then the best 4 predictor model (say $X_1,\ldots,X_4$). For the outcome Total Crime, the authors report the following coefficients of multiple determination.

8

$R^2(X_1,\ldots,X_{25}) = .913$ $\quad R^2(X_1,\ldots,X_4) = .774$ $\quad$ Compute $R^2(X_5,\ldots,X_{25} \mid X_1,\ldots,X_4)$

$$R^2(X_5,\ldots X_{25} \mid X_1,\ldots,X_4) = \frac{R^2(X_1,\ldots,X_{25}) - R^2(X_1,\ldots,X_4)}{1 - R^2(X_1,\ldots,X_4)}$$

$$= \frac{.913 - .774}{1 - .774} = \frac{.139}{.226} = .615$$

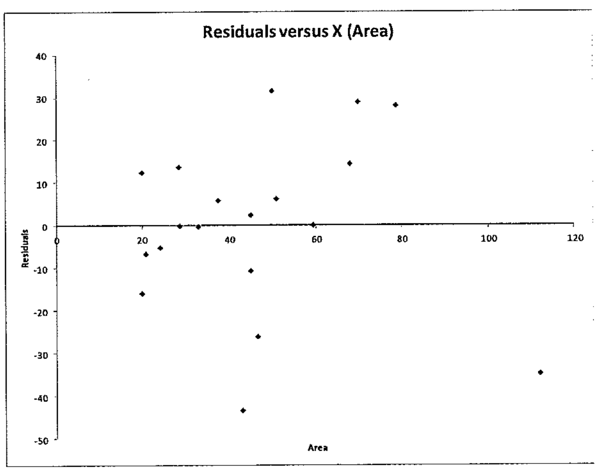
Q.6. A regression model is fit, relating energy consumption (Y) to total area (X) for a sample of $n = 19$ luxury hotels in Hainan Province, China. The Analysis of Variance for the simple linear regression model is given below.

ANOVA					
	df	SS	MS	F	gnificance
Regressio	1	25521.76	25521.76	57.75	0.0000
Residual	17	7512.94	441.94		
Total	18	33034.70			

p.6.a. A plot of the residuals versus area is given below. It demonstrates which possible violations of assumptions (circle all that apply).

2

Non-normal Errors $\quad$ Unequal Variance $\quad$ Serial Correlation of Errors $\quad$ Non-linear Relation between Y and X



$$X^2_{BP} = \frac{SSR^*/2}{\left(\frac{SSE}{n}\right)^2} = \frac{1239379/2}{\left(\frac{7512.94}{19}\right)^2}$$

$$= \frac{619689.5}{156355.3} = 3.963$$

p.6.b. A second regression model is fit, relating the squared residuals (Y) to area (X). Conduct the Breusch-Pagan test to test whether the equal variance assumption is reasonable. The sums of squares are given below.

ANOVA		
	df	SS
Regressio	1	1239379
Residual	17	3871645
Total	18	5111024

$X^2_{.05,1} = 3.841$

8

3 $\quad$ 2

Test Statistic: $X^2_{BP} = 3.963$ $\quad$ Rejection Region: $X^2_{BP} \geq 3.841$ $\quad$ P-value > or < .05

23

Q.7. A regression model was fit based on a sample of n=117 Black Holes. The response was Bolometric Luminosity (Y), with predictors: Black Hole Mass (X_1) and Black Hole Type ($X_2 = 1$ if Radio Quiet Quasar (RQQ), 0 if Radio Loud Quasar (RLQ)), and a cross-product term to allow for a possible interaction between Mass and Type. Note that there were 20 RQQ and 97 RLQ Black Holes.

Model 1: $\hat{Y}_i = 39.356 + 0.791X_{i1} + 2.047X_{i2} - 0.257X_{i1}X_{i2}$ $\quad SSE_1 = 30.964$ $\quad df_E^1 = 117-4 = 113$

Model 2: $\hat{Y}_i = 39.453 + 0.780X_{i1}$ $\quad SSE_2 = 31.187$ $\quad df_E^2 = 117-2 = 115$

p.7.a. Based on Model 1, give the fitted equations relating Bolometric Luminosity to Mass, seperately by Quasar Type.

$39.356 + 2.047 = 41.403 \qquad 0.791 - 0.257 = 0.534$

5 5

RQQ: $\hat{Y} = 41.403 + 0.534X_1$ RLQ $39.356 + 0.791X_1$

P.7.b. Test whether the true relationship between Bolometric Luminosity and Mass is the same for RQQs and RLQs.

H_0: $\beta_2 = \beta_3 = 0$ $\qquad H_A$: β_2 and/or $\beta_3 \neq 0$ 3

$$T.S.\ F_{obs} = \frac{\frac{31.187-30.964}{115-113}}{\frac{30.964}{113}} = \frac{.1115}{.2740} = 0.407$$

$F_{.05,2,113} \approx 3.077$

10 3 2

Test Statistic: $F = 0.407$ Rejection Region: $F \geq 3.077$ P-value > or < .05

p.7.c. When the regressions are fit seperately, the fitted equations are the same as you should have in part p.7.a. The residual variances (MSE's) for the models are: RQQ: $s_Q^2 = 0.088$ RLQ: $s_L^2 = 0.309$. Use Bartlett's test to test whether the true variances are equal.

$$B = \frac{1}{C}\left[\nu \ln(MSE) - \sum_{i=1}^{t} \nu_i \ln(s_i^2)\right] \qquad C = 1 + \frac{1}{3(t-1)}\left[\sum_{i=1}^{t} \nu_i^{-1} - \nu^{-1}\right] = 1.01$$

$N = 113$

$N_1 = 20-2 = 18$

$N_2 = 97-2 = 95$

$113 \ln(.2740) - (18 \ln(0.088) + 95 \ln(0.309))$

$= (-146.29) - [(-43.74) + (-111.57)] = (-146.29) + 155.31 = 9.02$

$B = \frac{9.02}{1.01} = 8.931 \qquad \chi^2_{.05,2-1} = 3.841$

Test Statistic: $B = 8.931$ Rejection Region: $B \geq 3.841$ P-value > or < .05

8 3 2

~~36~~ 41

Q.8. A random sample of n = 15 Bollywood movies was obtained, and a simple linear regression was fit relating the log Revenues (Y) the log Budget (X) ~~in international matches~~. Movie 15 was **Sultan** with log Revenues $Y_{15} = 5.705$ and log Budget $X_{15} = 4.500$. Regression models were fit with and without **Sultan**, and results are given below.

X'X			X'Y		X'X_(15)			X'Y_(15)	
15.000	55.474		53.796		14.000	50.974		48.091	
55.474	214.736		213.087		50.974	194.488		187.417	
INV(X'X)			Beta-hat		INV((X'X)_(15))			Beta-hat_(15)	
1.495	-0.386		-1.871		1.563	-0.410		-1.609	
-0.386	0.104		1.476		-0.410	0.112		1.385	
Y'Y	Y'PY				Y'Y_15	Y'PY_15			
218.04	213.79				185.50	182.26			

p.8.a. Obtain SSE and MSE and the estimated error standard deviation for each model.

$SSE_1 = 218.04 - 213.79 = 4.25$ $\quad df_E^1 = 15-2 = 13$ $\quad MSE_1 = 0.327$ $\quad S_1 = 0.572$

$SSE_2 = 185.50 - 182.26 = 3.24$ $\quad df_E^2 = 14-2 = 12$ $\quad MSE_2 = 0.270$ $\quad S_2 = 0.520$

$SSE_1 =$ 4.25 (4) $\quad MSE_1 =$ 0.327 (2) $\quad s =$ 0.572 (2)

$SSE_2 =$ 3.24 (4) $\quad MSE_2 =$ 0.270 (2) $\quad s_{(15)} =$ 0.520 (2)

p.8.b. Obtain the fitted value for **Sultan** for each model and its leverage value based on the full data set.

$\hat{Y}_1 = -1.871 + 1.476(4.5) = 4.771$

$\hat{Y}_2 = -1.609 + 1.385(4.5) = 4.624$

$$[1 \quad 4.5]\begin{bmatrix} 1.495 & -0.386 \\ -0.386 & 0.104 \end{bmatrix}\begin{bmatrix} 1 \\ 4.5 \end{bmatrix} = [-0.242 \quad 0.082]\begin{bmatrix} 1 \\ 4.5 \end{bmatrix} = 0.127$$

(4) (4) (6)

Fitted$_1$ = 4.771 $\quad$ Fitted$_2$ = 4.624 $\quad$ Leverage = $0.127 = v_{15,15}$

p.8.c. . Compute $DFFITS_{15}$ for **Sultan**. Note that DFFITS makes use of $s_{(15)}$ to estimate σ .

$$DFFITS_{15} = \frac{4.771 - 4.624}{0.520\sqrt{0.127}} = \frac{0.147}{0.185} = 0.795$$

$V\{\hat{Y}\} = \sigma^2 P$

$\Rightarrow V\{\hat{Y}_{15}\} = \sigma^2 v_{15,15}$

$DFFITS_{15} =$ 0.795 (6)

36